Word Recognition in Different Semantic Context

Research Thesis

Presented in partial fulfillment of the requirements for graduation
with research distinction in the undergraduate colleges of The Ohio State University

by

Shuchun Lai

The Ohio State University

April 2019

Project Advisor: Professor Mark M. Pitt, Department of Psychology

Abstract

Previous research found that people are more accurate recognizing spoken words in auditory scenes when the distracting sounds are nonspeech compared to speech (Toro, Sinnett, & Soto-Faraco, 2005). Working memory is a type of short-term memory; implying the ability to remember information for a brief period of time (Miyake & Shah, 1999). Working memory capacity (WMC) is the measurement of working memory, representing the capacity limit of information processed in working memory (Conway, et al., 2005). Kane and Engle (2000) found that variation in WMC is related to auditory selective attention. People with higher WMC show less disruption in a distractive listening task than those with lower WMC. Building on this work, Daly, Szostak, and Pitt (in preparation) have found that word recognition is more accurate for those with high than low WMC.

The present study brings these two lines of work together to ask whether WMC's influence on spoken word recognition differs for different distracting environments (speech vs. nonspeech). We tested the participants' WMC and have them do a distracting listening task where speech and nonspeech distractors are presented along with words to be recognized. For example, a sentence, e.g., *The ~~w~~ing had an exquisite set of feathers*, would be played along with a background distractor, and an extra noise mask is played at the onset of "w". The participants were asked to report what they hear the ambiguous word to be.

Past results lead to the prediction that when the distractor is speech, people with higher WMC will perform much better than people with lower WMC, and when the distractor is nonspeech, the performance difference between individuals with different WMC will greatly diminish. Our results weren't able to support our hypothesis, that no differences in performance in spoken-word recognition for people with different WMC and no differences in performance in different distractor types. The present study contributes to our understanding of the relationship between WMC and spoken word recognition in different noisy environments.

Word Recognition in Different Semantic Context

An active conversation involves the conveyance of a message which the recipient must understand. However, listeners often experience distracted speech, mainly whenever there exists ambiguity or lack of clarity that causes interferences. The common causes of distractions include noise from people talking in the background, sounds of street users, and visual distractors like passing people or vehicles, to name a few. These distractors all affect the effectiveness of speech recognition.

Selective attention is particularly important to allow the listener to utilize a filtering process to ensure the important signal is not disregarded (Dupoux, Kouider & Mehler, 2003). The most fortunate thing is that listeners can exploit the messenger's other means of delivering information, including semantic, prosodic, or gestural recognition to comprehend the message. Mattys, Carroll, Li, and Chan (2010) argued that these methods of acquiring information play a critical role as they ensure that attended-to-speech signal is processed during unintended distractions. When distractions are presented with the speech signals, abstract information such as the lexical or contextual information is more likely to be used to aid processing compared to lower level information such as the acoustic features of the signal (Mattys et al., 2010, Mattys & Wiget, 2011). Therefore, word recognition is negatively affected in a noisy environment to some extent.

It has been noted earlier that a noisy environment results in a distracted means of speech processing, but it would be even harder for the messenger to convey an intended message to the listener with the presence of a second speaker (e.g., Kozou et al., 2005; Mattys et al., 2010). Therefore, it requires individuals to have the ability to integrate distracting speech from the second talker with the main message without affecting the latter in distracting listening information. In the study of Mattys et al. (2010), they sought to examine what information individuals depend on when they hear speaking alone verses hearing speaking in the presence of a secondary talker. That is, how the participants' dependence on the phonologic-acoustic features and semantic-lexical features differs with the presence or absence of a secondary talker. In their study, the semantic-lexical features were controlled by changing the phrase from a real word phrase, *mild option*, to a pseudoword phrase, *mile doption*, and the second word in the phrase became a piece of potentially biasing information. Mattys et al. (2010) found that when the primary speaker is presented alone, listeners entirely relied on the acoustic-phonetic features, as opposed to semantic-lexical information in the signal. However, this was not the case when the main speech is given simultaneously, with a second speaker as listeners could depend more on lexical knowledge compare to the acoustical signals of information.

Previous research found that the semantic-lexical information had a profound impact when a listener faces speech distraction (Mattys et al., 2010). However, the semantic-contextual information was also found to be influential in distractive listening (Daly, Szostak, & Pitt, in preparation). That is, individuals' word recognition would not only be influenced by the words right next to each other, but other contextual information in the sentence. Research indicates the existence of biasing information using the possibility of an unclear segment which a listener can integrate to produce a meaningful sentence (Szostak & Pitt, 2013; Connine, Blasko &Titone 1993).

). For example, if a target speaker says, *"-ing* had multiple feathers," a listener might comprehend the *–ing* as *"wing."* In their study, for participants to be influenced by the whole sentential context of a word, their memory of the previous word must be maintained until the presence of later biasing information. Therefore, the more ambiguity of the memory of the previous word may result in a higher tendency to be biased by the latter word.

It has been noted that a noisy environment is significantly influential to speech processing, but the type of noise distraction may also be influential (e.g., Kozou et al., 2005). Kozou et al. (2005) observed that noise exposure principally changes the hearing threshold. The type of noises they used including noise of bubbles, industrial sounds, traffic, and wide band noise (Kozou et al., 2005). The noises decrease one's ability in word-processing, partly because of the reduction in their capacity to extract information from the speech. Most studies delve into auditory processing and have failed to compare the effects of different types of noise on word-processing (Kozou et al., 2005).

Toro, Sinnett, and Soto-Faraco (2005) discussed two types of distractors: speech and non-speech. Speech distractors include people or things that interfere with word-processing on the side of the listener by intercepting message delivery from the speaker. For example, a second speaker may interrupt the speech by uttering words that counteract the intended information. Non-speech distractors include all types of information that might be influential during a conversation. For example, sounds of moving objects, such as vehicles or other building or road users, may cause noise that would distract the main speaker.

In presenting speech distractors compared to non-speech distractors, the speech distractors affect more of the ability for listeners to extract the intended information (Toro et al., 2005). The negative impact on speech processing becomes substantial when a speech distractor is presented. This may be due to the fact that the speech distractor is more acoustically similar to the target speech, and the speech distractor contains semantic information that may entangle with those of the target speech. The results led the conclusions that the speech distractors have a more negative impact on word-processing compared to non-speech distractor.

Working memory (WM) and working memory capacity (WMC) also play an essential role in speech processing. WM is usually also referred as short-term memory that enables people to keep information in a short-span for cognitive process. According to Conway et al. (2005), WM is central in psychology, with the ability to perform tasks like counting, operation, and reading. In other words, WM is a multicomponent system that maintains information in continuous processing or distraction during message delivery. WMC is the measure of WM, which means the capacity an individual portrays when processing information in working memory (Conway et al., 2005).

According to Conway et al. (2005), WMC has two components: domain-general and domain-specific. The domain-general, also known as the executive control, is the ability of the listener to keep information in the active memory while simultaneously processing information due to interference. On the other hand, the domain-specific component is comprised of aspects that lead to the final storage of the information. These aspects include rehearsal and strategies that will improve one's ability to store the message (Conway et al., 2005, p.773). The use of a multi-domain system enables listeners to actively maintain information utilizing supplementary assistance.

There are exciting findings from the previous studies regarding executive attention and domain-specific elements of WMC. Executive attention predicts performance in various cognitive tasks, including those that are about language processing (DeCaro et al., 2008; Engle, 2002; Conway et al., 2001; Lustig et al., 2001; Otten & Van Berkum, 2009; Rai et al., 2010). For example, Conway et al. (2001) incorporated the names of the participants in the speech of a second speaker to measure the extent to which the WMC level could affect the ability of the listeners to detect their names when mentioned by the distractor. They found that individuals with low WMC recognize their names in an irrelevant message, which authors interpreted as their inability to block an unintended message. On the contrary, people with high WMC were less likely to report hearing their names. Furthermore, the authors affirmed their premise when they noted that the both the low and high WMC participants tended to not hear their own names. This suggests that the semantic information of the unintended message plays an important role in individuals' ability to block the irrelevant information. The scenario shows that individuals with lower WMC reach their capacity early compared to those with higher WMC; which explains their inability to filter or inhibit irrelevant information.

The growing research demonstrates the roles of WMC in speech processing in a distracted environment. The study by Conway et al. (2005) described how attention capacity could affect semantic information during speech processing. It is through attentional capacity that a listener could actively filter and discard unrelated speech to enhance one's ability to process the intended message. Conway et al. (2005) used this construct to build the WMC framework that could measure a listener's ability to handle speech in an interrupted environment. Kane and Engle (2000) found a correlation between WMC and auditory selective attention; whereby they demonstrated that individuals with higher WMC scores showcased less disruption in a distracted listening environment. They found a reverse situation among listeners with low WMC, where they were negatively influenced by the distraction to a larger extent.

Toro et al. (2005) concluded that speech processing demands a significant amount of attention. It is at this point that the WMC plays a critical role in processing speech. According to Matzen and Benjamin (2009), memory performance relies on the ability of listeners to remember or trace words for biasing information to exert sufficient impact on word identification. Therefore, it means that a processor must overcome the secondary speaker's speech to maintain a certain level of the memory trace. A listener with higher WMC has exceptional abilities in maintaining information in their WM during speech processing in noisy environment, compared to those with lower WMC (Conway, Cowan & Bunting, 2001. pp. 334). This suggests that maybe the amount of information maintained in individuals' WM is related to their selective listening ability. Kane and Engle (2000, pp. 341-342) made similar observations, whereby they found supporting evidence that listeners with high WMC demonstrate less disruption in a distracting environment, compared to their peers with lower WMC. Therefore, for listeners with low WMC, it is possible that when faced with a distracting situation with the possibility of encountering difficulties, then they may depend highly on the semantic information, as opposed to the phonetic representation of the primary word. In conclusion, a noisy environment tends to significantly interfere with speech processing in both speech and non-speech distractions, although this largely depends on the WMC of the recipient.

In reviewing the literature, we can see that some research found that the variation in WMC is related to spoken word recognition (e.g. Kane & Engle, 2000), especially that people with high WMC have more advantages in word recognition (e.g. Daly et al., in preparation; Conway et al., 2001). However, there remains space for further investigation of whether WMC's influence on spoken word recognition differs for different distracting environments (speech vs. nonspeech).

The present study investigated if the biasing information at the end of a sentence would influence recognizing previous words in the sentence. For example, for the sentence "*The wing had an exquisite set of diamond",* if changing the last word in the sentence *diamond* would influence people's perception of the second word *wing*. Further, we studied if people with different WMCs could have different performances in spoken-word processing in the presence of speech distractor relative to nonspeech. The present study was expected to show that individuals with low WMC identify ambiguous words with predominant dependence on biasing information. We predicted that when the distractor is speech, people with higher WMC would perform much better than people with lower WMC; and when the distractor is non-speech, the performance difference between individuals with different WMCs would greatly diminish. The measurement for the performance is an extension of the measurements used in Szostak and Pitt (2013), and Connine, Blasko, and Titone (1993).

If our finding is consistent with the prediction, it would suggest that WMC influences spoken word identification much more when the distractor is speech, compared to non-speech.

Method

Participants

The study involved a sample of 24 participants who had obtained course credits in an introduction to psychology course in the Ohio State University for their willingness to partake in the experiment. All participants were self-reported native American English speakers with normal hearing capacities.

Stimuli

In the present study, we adopted the sound stimuli from experiment 1 and experiment 4 by Daly, Szostak, and Pitt (in preparation) and Szostak and Pitt (2013). The study selected 24 target words for the main experiment and 20 target words for the practice block, each a monosyllabic noun with at least one possible rhyme competitor. The target words were used to construct a pair of sentences. There was a noise mask presented onset of the first phoneme, such as a noise mask was presented at the onset of “w” for the word “*wing”.* Throughout the experiment, the target word was the second word. Then, the final word in each pair of the sentence was biased either toward or against the target word. The congruent condition is when the final word was biased toward the target. For example, “*The ~~w~~ing had an exquisite set of feathers*.” The incongruent condition is when the final word of the sentence was biased against the target word. For example, “*The ~~w~~ing had an exquisite set of diamonds*,” whereby a ring would be semantically linked to diamonds. The study ensured there are six to eight syllables between the target word and the biasing word. In this way, we would be able to explore the contextual effect for words that appeared in a great distance in the sentence, and thus, investigate if the ability to hold information in WM for a few seconds is related to the performance in spoken-word recognition. The stimuli were recorded by the second author, a female native speaker of American English. The final word was spliced from the sentence. Two copies were retained from the remainder of the sentence and saved. A congruent biasing word from another token of the same sentence was then placed into one of the remaining fragmented sentences. For example, “*The ~~w~~ing had an exquisite set of + feathers*,” compared to an incongruent biasing word from the sentence, “*The ~~w~~ing had an exquisite set of + diamond*.” Using the same copy of the sentence with the addition of different tokens of the last word ensured the two sentences to be identical in the largest extent.

The next step involved locating the onset phoneme of the target word auditorily and visually with the help of a spectrogram. Subsequently, the addition of noise to the signal was performed using the Samuel (1987) technique. The method was appropriate as it ensured sound had the same amplitude envelope as the masked phoneme, and it let the noise have a syllable-like quality. Twelve females from Buckeye Corpus of Conversational Speech produced the distracting speech, which served as the stimuli for the secondary talkers in the right ear (Daly et al., in preparation; Pitt, Johnson, Hume, Kiesling, & Raymond, 2005). Primary sentences recorded by the second female author were paired with those from secondary talkers from the Buckeye Corpus of Conversational Speech (Pitt et al., 2005; Daly et al., in preparation). Throughout the experiment, it was made sure that the target talker’s sentence appeared at the left ear and the counterpart at the right ear. The target sentence was always presented in the left ear across participants because previous research found that the majority of people who are right-handed have a right-ear advantage for hearing speech. This means they were more likely to be distracted to hear primary sentences presented in the left ear (e.g., Foundas, et al, 2006).

A corpus of 208 familiar sounds (for example, animal and household sounds) helped in constructing non-speech distractors. These sounds were adopted from Daly et al. (in preparation), which were also similar to stimuli used in Toro et al. (2005). Each sound ranges between 100 msec to 1,000 msec in length, and each of them can be identified with the source that can produce the sound. To avoid confusing sound and speech, sounds like singing, coughing, and sneezing to name a few were excluded. Subsequently, a group of sound files were joined without intervening silence. The pairing of sentences in different ears aimed to ensure that the resulting sound chain was within a close span to the total duration. We used the sound chains to be more similar to the daily speaking situations. That is, the sound chains were able to have more variability in timbre, and this brought more challenges to the listener. Upon the construction of sound chains, the mean amplitude was set at 70 dB, the same as that of the secondary talker. Then, the sound chains were combined with target talker sentences.

Procedure

There were two parts to the experiment. Which task to do first was counterbalanced across participants. In both of the tasks, each participant stayed individually in a sound-attenuated room.

Working Memory task. To measure WMC, the study used the Updating task in a modified form compared to the one utilized in Miyake et al. (2000). The Updating task was selected because Wilhen, Hildebrandt, and Oberauer (2013) found positive correlations between the updating collection of tasks and the complex span tasks, and Schmiedek, Lövdén, and Lindenberger (2014) confirmed the finding. Since the Operational Span (Ospan) task is historically one of the most frequently used tasks shown in an extensive amount of literature (e.g., Amir & Bomyea, 2011; Aslan & Bäuml, 2011; Redick et al., 2012), we did a prior study comparing the Updating Working Memory task with the Ospan task. We found that the Updating task has good variability in WMC scores, and it also has a good correlation with the Ospan task ($r=0.49$).

Participants would see a series of boxes presented horizontally on the center of the screen, and need to remember the digits shown in the boxes. Only one number would appear at a time, and each box eventually would have at least one digit shown in it. However, participants would be required to remember the last digit that appeared in each box, forgetting all other digits shown previously in that box. At the end of each test, participants would type in the last digit they remembered into each of the highlighted boxes on the screen, and the sequence of which box was highlighted, i.e., the sequence of which box to answer was randomly simulated. The accuracy of their response will be the measurement of their WMC. The higher the accuracy, the higher their WMC.

In this task, we used digits from 0-9 for participants to remember, and we presented each digit randomly in one of the boxes laid out on the screen. The height of the digit size was 1/10 of the monitors' height, and the width and the height of the box was 1/5 of the monitors' height. The number of boxes being presented varied from 3-7, even though only the responses in box sizes 5-7 were used in scoring WMC. Box numbers 3 and 4 were not used in scoring WMC, were designed and used so that the participants did not feel the task to be overly challenging. There are 3 practice trials followed with 19 experimental trials, including 5 trials each for box sizes 5-7 and 2 trials each for box sizes 3-4. The possible number of updating steps of each box varied from 2-6. Different box sizes can differentiate individual differences in the updating domain of WM. The stimulus duration for each digit is 1.6 seconds, and the internal stimulus interval (ISI), i.e. the interval between last digit disappearing and next digit appearing, was 0.5 seconds.

Dichotic Listening Task. The participants did the tasks in an individual sound dampened room. They were exposed to stimuli through headphones. Before experiencing the experimental blocks, participants first did a practice block that contains 20 practice trials. The practice trials included stimuli that were different from the experimental block, including both the congruent and incongruent conditions, plus the speech and non-speech conditions. There were 4 blocks of experiments in total, two of those blocks include only stimuli when the distraction in the right ear was speech, and the other two blocks included only stimuli when the distraction was non-speech. The question of which two blocks to present first was counterbalanced. Each block contained 24 trials, including both congruent and incongruent conditions. The congruent and incongruent trials were pseudo-randomized in each block. The experiment ensured that no target word could be repeated within the same block. This means that the two semantic conditions on congruency for the target word only appeared in two different blocks. Therefore, each block included 24 distinct target sentences with different target words.

Furthermore, after the first two experimental blocks, the program would stop for 30 seconds so that participants could take a break, and then they would do 3 more practice trials as they did at the beginning, before going to the next two experimental blocks. However, the 3 practice trials only include either the speech or non-speech condition, and which condition was presented was determined by the condition for the last two blocks. For example, if the last two blocks were the ones including the speech condition, the 3 practice trials would also all be in the speech conditions.

In each trial, participants were participated to a single stimulus item, whereby the participant could type what he/she heard the target word was after the talkers had finished speaking. At the beginning of the experiment, the participants were encouraged to type the word they actually heard rather than what they thought they were supposed to hear. They were also guided toward the understanding that the target word was always the second word in a sentence. Following the typing response in each trial, participants encountered a question of whether the previous sentence made sense or not. Then they pressed the “s” key on the keyboard to confirm if they judged the sentence to have made sense, or “n” if otherwise. The sensibility questions were designed to make sure that the participants are actively paying attention to the entire sentences so that they were trying to comprehend the meaning of the sentence. The computer then moves to the next trial. It generally took a total of 25-35 minutes to complete the experiment.

In this task, we used two measurements of participants’ performances: 1) the proportion of correct answers, and 2) the proportion of being biased. A response was taken to be correct if the reported word is consistent with the target word. Such as reporting "wing" when hearing either the sentence, "the wing had an exquisite set of feathers" or, "the wing had an exquisite set of diamonds." However, a sentence would be taken to be biased only if the reported word was semantically related to the last word in the sentence. The measurement of biasedness has been adopted from Daly et al. (in preparation), and it was designed to capture the tendencies of how people are influenced by the semantic context. A response can be scored as both accurate and biased, and this would usually be the case when participants reported hearing "wing", and when the last word was "feathers", (i.e. report target word in the congruent condition). The participant would be scored as 1 if they correctly reported the target word or scored as 0 if not. We omitted the trials in our analysis when participants accidentally reported other words rather than the second word that was presented, such as reporting "had" when they hear "the wing had an exquisite set of features".

Results

Our hypothesis is that people with higher WMC would show higher accuracy and less likely to be biased when the distractor is speech compared to those with low WMC, and we expect this differences in performance would diminish when the distractor is nonspeech. We also expected an overall less accuracy and more bias of recognizing the word when the distractor is speech.

From the experiment, the overall proportion of correct responses was 0.57, and proportion of bias was 0.56. However, both of these two performance measurements differ significantly under different contextual congruency conditions. The results are shown in *Figure 1*. The participants' average proportion of correct response (accuracy) in the congruent condition is higher than the incongruent condition ($t(23) = 6.34$; $p<0.001$), and the participants' average proportion of being biased is even higher in the congruent condition compared to the incongruent condition ($t(23) = 14.78$; $p<0.001$). It makes sense that participants had more biased responses in the congruent condition. Participants were scored as biased if their responded words were semantically related to the last words in the sentences, and thus they were almost always scored as biased if they accurately reported the target words in the congruent condition. However, since the participants were not scored as biased when they accurately hear the target word in the incongruent condition, it is unsurprised that the proportion of biasness is higher in the congruent conditions.

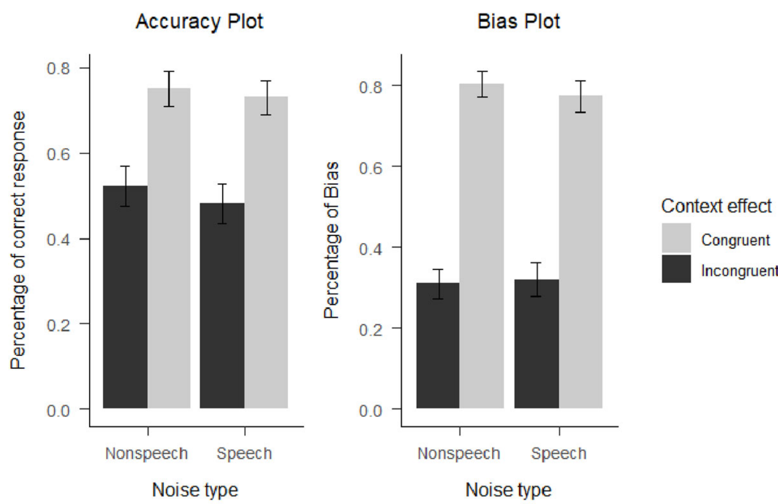

Figure 1

The results of the comparison between performance under the speech and non-speech conditions is found to be insignificant for measurement of accuracy ($t(23)=0.68$; $p=0.50$) and proportion of bias ($t(23)=0.16$; $p = 0.87$). This result is surprisingly inconsistent with the suggestion from the previous studies, in that we didn’t find a significant increase of accurate responses or decrease of bias responses under the nonspeech distractor condition (e.g. Daly el al.).

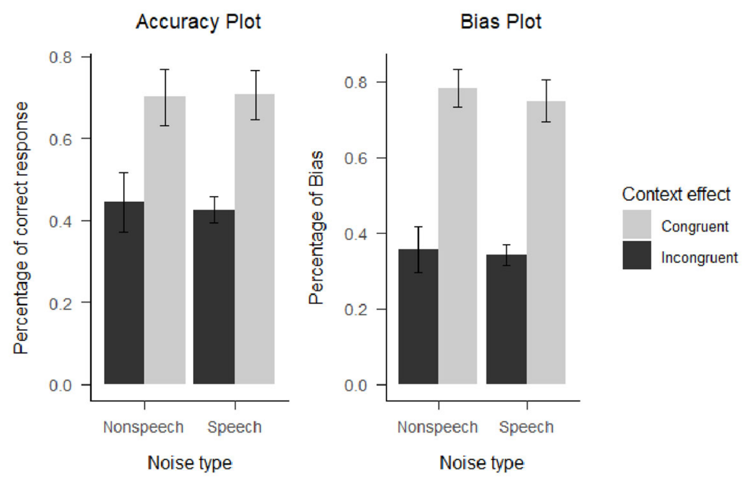

Figure 2

To probe the possibility that practice effects were masking differences due to distractor type, we conducted a between-participants analysis where only the first two blocks for each participant were included in the analysis. In this way, each participant was only exposed to one type of distractor in this analysis. The results are shown in Figure 2. The conclusions were similar to the original analysis. The participants' average proportion of correct response is higher in the congruent condition compared to the incongruent condition ($t(11) = 5.42$; $p<0.001$), and the participants' average proportion of biased response is higher in the congruent condition compared to the incongruent condition ($t(11) = 9.81$; $p<0.001$). The results of the comparison between performance under the speech and non-speech conditions is still found to be insignificant for measurement of accuracy ($t(11)=0.10$; $p=0.91$) and proportion of bias ($t(11)=0.32$; $p = 0.74$).

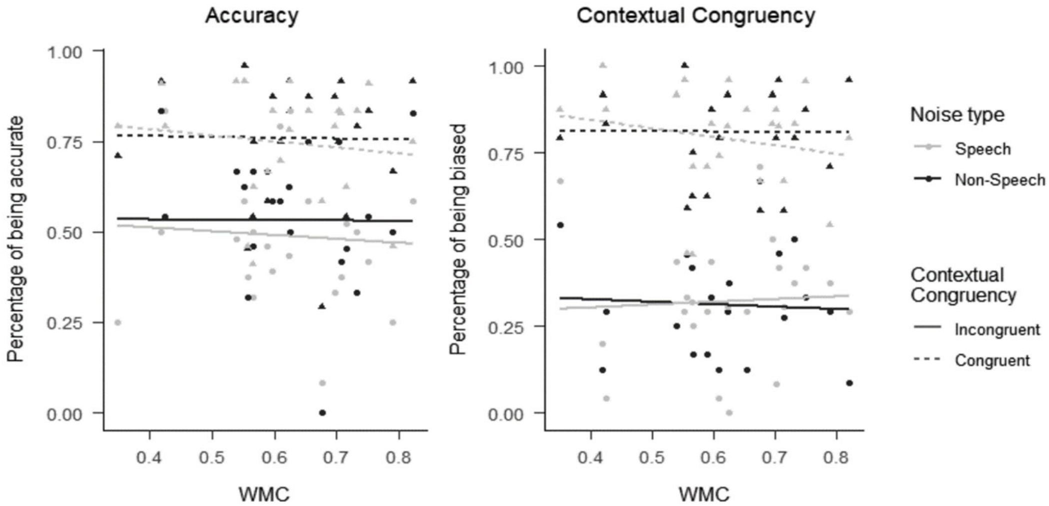

Figure 3

The influence of WMC is inconclusive in the current study for both measurements (accuracy and bias) of performance in the selective listing task. The results are shown in Figure 3. In the incongruent condition, the correlation between WMC and accuracy is -0.064 when noise type is speech, and -0.006 when noise type is nonspeech. For the measurement of bias, in the congruent condition, the correlation between WMC and bias is 0.053 when noise type is speech, and -0.052 when noise type is nonspeech. In the incongruent condition, the correlation between WMC and accuracy is -0.064 when noise type is speech, and -0.006 when noise type is nonspeech. For the measurement of bias, in the congruent condition, the correlation between WMC and bias is 0.053 when noise type is speech, and -0.052 when noise type is nonspeech. We treated participant, target words and speech (speech or nonspeech) as random factors, and WMC as fixed factors. We used the lme4 package in the *R statistical computing* environment to conduct logistic mixed-effects regression on the data of incongruent condition (Bates, Maechler, & Bolker, 2014). The mixed models did not detect significant fixed effect for WMC of accuracy in speech condition ($\beta = 0.429, SE = 1.54, z = 0.243, p = 0.808$) or nonspeech condition

(β=0.548 , SE=0.805 , z=0.681 , p= 0.496), or of bias response in speech condition (β=0.696 , SE=3.231 , z=0.215 , p= 0.829) or nonspeech condition (β=-1.546 , SE=2.777 , z= -0.557 , p= 0.578).

We also conducted between-participants analysis to find the association between WMC and performances in different contextual and noise type conditions. The results are shown in figure 4. The correlation between WMC and accuracy is 0.110 in non-speech and incongruent condition, and -0.178 in speech and incongruent condition; the correlation between WMC and proportion of being biased is -0.161 in non-speech and incongruent condition, and 0.355 in speech and incongruent condition. Similar mixed regression model was used to perform the between-group analysis. The mixed models did not detect significant fixed effect for WMC of accuracy in speech condition ($\beta = 0.375, SE = 1.54, z = 0.243, p = 0.808$) or nonspeech condition (β=0.223, SE=0.160, z=0.139, p=0.889), or of bias response in speech condition (β=0.862, SE=1.619, z=0.532, p=0.595) or nonspeech condition (β=-0.175, SE=1.490, z=-0.117, p=0.907).

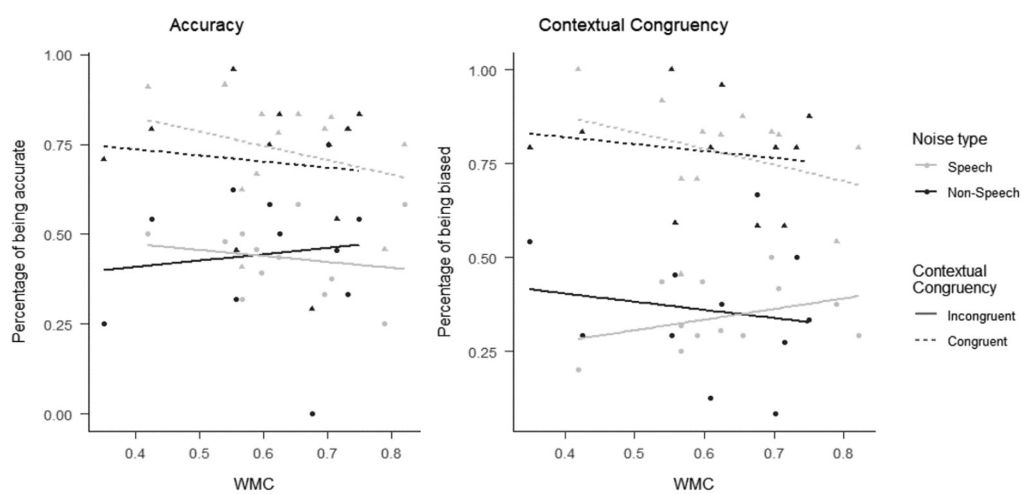

Figure 4

Because there did not seem to be any effects of WMC, we decided to look at the distribution of WMC scores. The average score for WMC was 0.6216, and the individuals' scores ranged from 0.35 to 0.82. The distribution of WMC seems is right skewed. The density distribution of WMC is shown in figure 5. This distribution of WMC shows that we had more participants with WMC around the middle range, but we had less participants who have extreme WMC scores.

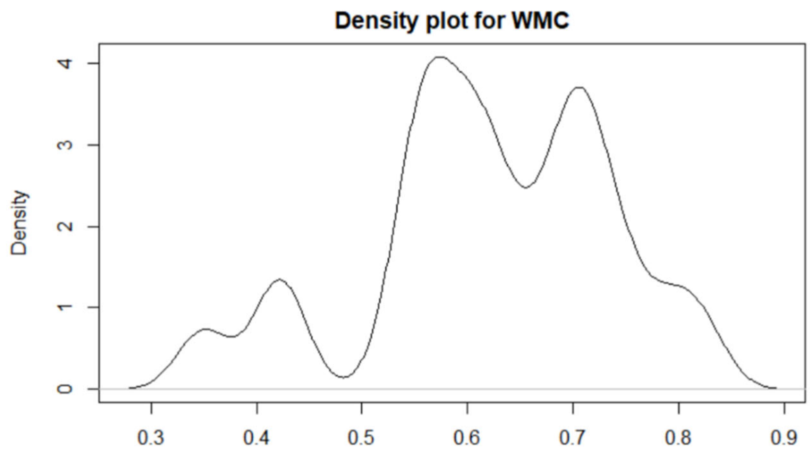

Figure 5

Discussion

The current study was designed to find out whether WMC's influence on spoken word recognition differs for different distracting environments (speech vs. nonspeech). Our results regarding this question do not support the hypothesis that WMC influence the spoken word recognition in different distracting environment.. Previous research suggests that people having a high WMC showed greater accuracy in word recognition; while those with low WMC were biased and the accurate word identification was a challenge to them (e.g. Daly, et al., in preparation). However, in the current study, it's not significant if WMC plays a role in predicting different performances in various conditions. According to Conway et al. (2005), when an individual has high WMC, there should be more executive attention available of being used.

When the distractor is another speaker, people with low WMC have less executive attention to distinguish the noise from the attended speech. The attention will help in word identification and boost the level of accuracy that is obtained through the experiment. However, the results from the present study were not able to verify this prediction..

There are couple of limitations of this study that may result in our insignificant results on the effect of WMC in performance of word recognition. First, we think the fact that most of the participants in the present study had moderate to high WMC may influenced our results on finding an effect of WMC. It would mean that we were not able to capture the effect of WMC on word recognition due to the fact that we merely did not have data on participants who have low WMC. Second, the insignificant results may due to limits in our original within-participant design. In our original design, the participants experienced different distractor types when they went from the first two blocks to the last two blocks. The target words were the same for both distractor type, but the distractor type changed only after the first two blocks. Therefore, participants might learn to perform better after doing the first two blocks, or the they might merely get tired to do the last two blocks. All these proposed reasons would contaminate our results in the last two blocks and therefore influenced our overall findings. It was because of this reason that we did the between-group analysis. Even though the findings in the between-group analysis is still statistically insignificant, the influenced of WMC seem to be Finally, for the partial of the experiment, we used an flawed version of WMC task that only have 15 experimental trials while the correct version have 19 experimental trials. In the flawed version of WMC task, the last 4 trails always weren’t presented to the participants. This could be an important factor that influenced our results.

The obtained results established an inconsistent outcome on the effect of distractor type (speech vs. nonspeech) compared with the other previous studies (e.g. Van-Tasell and Yanz, 1987; Kozou et al., 2005). It was expected that people would show much lower accuracy in word recognition in the presence of speech as background noise, compared to non-speech. Further, people should show more biased responses when the distraction is speech compared to non-speech. When the distraction is speech, the semantic information may also bias recognition for the primary speech. Furthermore, because we expect it to be more distracting when the distractor is speech compared to nonspeech, people may end up replying more on the context when distracted by speech.

Accurately recognizing the target word was thought to be hard, also because the second speaker would have more similar phonetic features compared to when the distraction is non-speech. It was quite surprising to find an inconclusive result in the study on the effect of noise type, especially because many participants verbally reported the task is harder when the distraction is speech compared to non-speech. Compared to the results in Daly et al. (in preparation), where they found a more significant difference between different noise types by comparing data across two experiments, the current study found little significant difference. Considering that the current sound stimuli was adopted from Daly et al. (in preparation), the inconsistency of findings between this study and their study is even surprising.

In conclusion, we found the effect of context on spoken word recognition, that people are more likely to be accurate in the congruent condition. We did not find an effect of distractor type or the influence of WMC on spoken word recognition. Considering the limitations of the study, we suggest further study on including more people in lower WMC span.

References

Amir, N., & Bomyea, J. (2011). Working memory capacity in generalized social phobia. *Journal of Abnormal Psychology, 120*, 504-509.

Aslan, A., & Bäuml, K. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 264-269.

Bates, D. M., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed– effects models using S4 classes (R package version 0.999999 – 0). Retrieved from http://CRAN.R-project.org/packagelme4

Conway, A., Cowan, N., & Bunting, M. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331-335. doi: 10.3758/bf03196169

Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786. doi: 10.3758/bf03196772

Daly, H., Szostak, C.M., & Pitt., M. (In preparation) *A behavioral investigation of the mechanisms underlying auditory selective attention*.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*, 193–210.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition, 107,* 284-294.

Dupoux, E., Kouider, S., & Mehler, J. (2003). Lexical access without attention? Explorations using dichotic priming. *Journal Of Experimental Psychology: Human Perception & Performance*, *29*(1), 172-184. doi: 10.1037//0096-1523.29.1.172

Engle, R. W. (2002). Working memory as executive attention. Current Directions in *Psychological Science, 11*, 19-23.

Foundas, A. L., Corey, D. M., Hurley, M. M., & Heilman, K. M. (2006). Verbal dichotic listening in right and left-handed adults: Laterality effects of directed attention. *Cortex, 42*, 79-86.

Kane, M., & Engle, R. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *26*(2), 336-358. doi: 10.1037/0278-7393.26.2.336

Kozou, H., Kujala, T., Shtyrov, Y., Toppila, E., Starck, J., Alku, P., & Näätänen, R. (2005). The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hearing Research*, *199*(1-2), 31-39. doi: 10.1016/j.heares.2004.07.010

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130*, 199-207.

Mattys, S., Carroll, L., Li, C., & Chan, S. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, *52*(11-12), 887-899. doi: 10.1016/j.specom.2010.01.005

Matzen, L., & Benjamin, A. (2009). Remembering words not presented in sentences: How study context changes patterns of false memories. *Memory & Cognition*, *37*(1), 52-64. doi: 10.3758/mc.37.1.52

Mattys, S. L. and Wiget, L. (2011). "Effects of cognitive load on speech recognition." J. Mem. Lang. 65, 145-160.

Miyake, A., & Shah, P. (Eds.). (1999). Models of working memory: Mechanisms of active maintenance and executive control. New York, NY, US: Cambridge University Press. Doi: /10.1017/CBO9781139174909

Miyake A., Friedman N. P., Emerson M. J., Witzki A. H., Howerter A., Wager T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology ,41*, 49–100.

Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research, 1291*, 92-101.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye Corpus of Conversational Speech: Labeling conventions and a test of transcriber reliability. *Speech Communication, 45*, 89-95.

Rai, M. K., Loschky, L. C., Harris, R. J., Peck, N. R., & Cook, L. G. (2010). Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. *Language Learning, 61*, 187-218.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*, 164-171.

Samuel, A. G. (1987). Lexical uniqueness on phonemic restoration. *Journal of Memory and Language, 26*, 36-56.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). Younger adults show long-term effects of cognitive training on broad cognitive abilities over 2 years. *Developmental Psychology, 50*(9), 2304-2310. doi:10.1037/a0037388

Szostak, C. M., & Pitt, M. A. (2013). The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, & Psychophysics*, *75*(7), 1533–1546. doi:10.3758/s13414-013-0492-3[

Toro, J., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25-B34. doi: 10.1016/j.cognition.2005.01.006

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in psychology, 4*, 433. doi:10.3389/fpsyg.2013.00433